Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
Available online at www.inia.es/sjar

Spanish Journal of Agricultural Research 2011 9(1), 156-165
ISSN: 1695-971-X
eISSN: 2171-9292

Genetic redundancy among durum wheat accessions as assessed by SSRs and endosperm proteins

M. Ruiz[1]*, E. Aguiriano[1], P. Giraldo[2] and J. M. Carrillo[2]

[1] *Centro de Recursos Fitogenéticos (CRF). Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA). Autovía de Aragón, km 36. Apdo. 1045. 28800 Alcalá de Henares (Madrid). Spain*
[2] *Departamento de Biotecnología. ETSI Agrónomos. Universidad Politécnica de Madrid (UPM). Ciudad Universitaria. 28040 Madrid. Spain*

Abstract

Reducing duplication in *ex-situ* collections is complicated and requires good quality genetic markers. This study was conducted to assess the value of endosperm proteins and SSRs for validation of potential duplicates and monitoring intra-accession variability. Fifty durum wheat (*Triticum turgidum* ssp. *durum*) accessions grouped in 23 potential duplicates, and previously characterised for 30 agro-morphological traits, were analysed for gliadin and high molecular weight glutenin (HMWG) subunit alleles, total protein, and 24 SSRs, covering a wide genome area. Similarity and dissimilarity matrices were generated based on protein and SSRs alleles. For heterogeneous accessions at gliadins the percent pattern homology (PH) between gliadin patterns and the Nei's coefficient of genetic identity (I) were computed. Eighteen duplicates identical for proteins showed none or less than 3 unshared SSRs alleles. For heterogeneous accessions PH and I values lower than 80 identified clearly off-types with more than 3 SSRs unshared. Only those biotypes differing in no more than one protein-coding locus were confirmed with SSRs. A good concordance among proteins, morphological traits, and SSR were detected. However, the discrepancy in similarity detected in some cases showed that it is advisable to evaluate redundancy through distinct approaches. The analysis in proteins together with SSRs data are very useful to identify duplicates, biotypes, close related genotypes, and contaminations.

Additional key words: cereals; duplicates; *ex-situ* collections; molecular data; prolamins.

Resumen

Evaluación de la redundancia genética con microsatellites y proteínas del endospermo en accesiones de trigo duro

Reducir la existencia de duplicados en las colecciones *ex-situ* es una tarea complicada que require el uso de buenos marcadores geneticos. En el presente trabajo, se evalúa el valor de las proteínas del endospermo y de los SSRs para la identificación de duplicados potenciales y el análisis de la variabilidad dentro de las accesiones. Se han seleccionado 50 accesiones de trigo duro (*Triticum turgidum* ssp. *durum*), agrupadas en 23 duplicados potenciales y previamente caracterizadas para 39 caracteres agro-morfológicos. Se ha analizado su composición en proteina total, alelos de gluteninas de alto peso molecular (HMWG) y 24 SSRs, cubriendo gran parte del genoma. Con los datos de los alelos de proteínas y SSRs se han generado matrices de similitud y disimilitud. En las accesiones heterogéneas para gliadinas se ha calculado el porcentaje de homología entre los patrones de gliadinas (PH) y el coeficiente de identidad genética de Nei (I). Los 18 duplicados idénticos en su composición en proteínas mostraron menos de 3 alelos SSRs diferentes. En las accesiones heterogéneas, los valores de PH e I menores de 80 identificaron a los individuos fuera de tipo, con más de 3 alelos SSRs diferentes. Solamente aquellos biotipos que diferían en no más de 1 locus proteíco se confirmaron con SSRs. Se ha detectado una buena concordancia entre los datos de proteínas, los SSRs y los caracteres agromorfológicos. Sin embargo, las discrepancias observadas en algunos casos avalan la necesidad de evaluar la redundancia mediante distintos marcadores. El análisis conjunto de los datos de proteínas y SSRs es muy útil en la identificación de duplicados, biotipos, genotipos cercanos y contaminaciones.

Palabras clave adicionales: cereales; collectiones *ex-situ;* datos moleculares; duplicados; prolaminas.

* Corresponding author: mruiz@inia.es
Received: 10-05-10; Accepted: 18-11-10.

Abbreviations: A-PAGE (acid polyacrylamide gel electrophoresis), d_{GP} (Goldstein and Pollock distance), d_{RW} (distance of Rogers as modified by Wright), HMWG (high molecular weight glutenin), I (Nei coefficient of genetic identity), PH (percent pattern homology), s_J (Jaccard's similarity index), SSR (short sequence repeat).

Introduction

In a broader sense, genebank duplication can be defined as accessions derived from the same original population, having alleles in common. Historical duplicate accessions, however, may diversify from the primary sample in genetic composition during maintenance in *ex-situ* collections (Hintum and Knüpffer, 1995). For these reasons duplicate entries would be expected to be equally different from each other and from their common original accession, and thus form a genetically homogeneous group (Lund *et al.,* 2003). But the amount of genetic diversity acceptable between duplicate entries is still not well-defined. Consequently, reducing duplication is more complicated than generally assumed and requires good quality genetic markers to estimate genetic similarities. The potential use of these markers depends on their polymorphism and repartition on the genome, avoiding redundant information. Speed, cost, and reproducibility determine their utility for genebank management.

In wheat, morphological and physiological characters which were traditionally used provide practical information to breeders, but they are not sufficient because of low polymorphism and variation under environment. Molecular markers at the protein level such as endosperm proteins (gliadins, glutenins, albumins, and globulins) are primary products of gene expression and can reveal small changes (*e.g.* mutations) inaccessible to visual examinations. The extensive heterogeneity of gliadin electrophoretic composition confers a high level of discrimination among wheat cultivars (Sapirstein and Bushuk, 1985; Metakovsky, 1991; Kudryavtsev *et al.,* 1996). Allele identification is also useful to distinguish biotypes from an off-types, or admixtures, in heterogeneous accessions (Metakovsky, 1991; Kudryavtsev *et al.,* 1996). The disadvantage of this approach is that gliadins are encoded by only six loci *(Gli-*) on the short arms of chromosomes of the first and sixth homoeologous groups (Payne *et al.,* 1982). Electrophoresis analysis of total endosperm protein, including albumins and globulins, and HMWG (high molecular weight glutenin) subunits allows a wider genome coverage because they are controlled by groups 1, 2, 3, 4, 5, 6, and 7 of wheat chromosomes (Fra-Mon *et al.,* 1984; Singh and Skerritt, 2001).

Nowadays, characterisation of germplasm by means of DNA fingerprinting techniques supplies a tool for a precise estimate of genetic diversity. In this context, SSRs are high polymorphic genetic markers with a uniform distribution in the wheat genome and are very useful for studying the variability of wheat germplasm (Röder *et al.,* 1998; Maccaferri *et al.,* 2003).

In previous research, we identified 106 potential duplicates of durum wheat accessions maintained at the National Plant Genetic Resources Centre (Spain) using passport data, 30 agro-morphological characters and gliadin banding patterns (Ruiz and Aguiriano, 2004). In some cases, potential duplicates similar in gliadin patterns differed in two or more agro-morphological characters indicating that more molecular markers, including SSRs, were required for resolving these cases. Furthermore, the identification of gliadin alleles would allow the computation of genetic distances between duplicate accessions and to identify admixtures in heterogeneous accessions.

In the present research, some of the duplicates previously studied have been analysed for gliadin and HMWG subunits alleles, total protein, and SSR alleles. The objective was to assess the value of endosperm proteins and SSRs for validation of potential duplicates, and monitoring the intra-accession variability.

Material and methods

Materials

Fifty durum wheat accessions grouped in 23 potential duplicates were selected from a previous study (Ruiz and Aguiriano, 2004). Some duplicates were similar in gliadin patterns but differed in several agromorphological traits, and others showed concordance between the two data sets. Potential duplicates (landraces or cultivars) had identical variety names and contained two or three accessions. One cultivar, Andalucía 344, was also included because of its high intra-accession gliadin variability.

Agro-morphological traits

The accessions were evaluated in previous research (Ruiz and Aguiriano, 2004) for 30 agro-mophological characters habitually used for wheat variety identification.

Endosperm protein analysis

As a minimum, ten grains per accession were analysed for intra-variety characterisation for gliadins, HMWG subunits, and total protein. Gliadins were extracted from half single seeds and fractionated in acid (pH 3.1) polyacrylamide gel electrophoresis (A-PAGE) according

to Lafiandra and Kasarda (1985). Identification of most of the *Gli*-alleles was performed in previous research (Aguiriano *et al.,* 2006) following the catalogue of Kudryavtsev *et al*. (1996). Electrophoresis of total protein and HMWG subunits was performed on sodium dodecyl sulphate polyacrylamide gels according to Payne *et al*. (1980). HMWG subunits alleles at *Glu-A1* and *Glu-B1* loci were identified following the nomenclature of Payne and Lawrence (1983).

SSR analyses

All the genotypes different in protein composition were examined with SSRs. Three genotypes (accessions Granja de Badajoz and Raspinegro) were not analysed because weak amplification products were produced. The same grain analysed for protein was germinated and DNA was isolated from fresh leaves following a protocol of Doyle and Doyle (1990). Twenty-three primer pairs, selected based on their independent genomic distribution, profile quality and polymorphism level, were used to amplify the A and B genomes. Primer sequences, reaction mixture, and PCR cycles were same as described by Röder *et al*. (1998). The forward primers of each primer pair were fluorescently labelled with 6-carboxyfluorescein (6-FAM), hexacloro-6-carboxyfluorescein (HEX), and tetrachloro-6-carboxyfluorescein (TET). The allele size was analysed in an ABI PRISM 310 Genetic Analyser. The method described by Ghosh *et al*. (1997) was followed to determine the integer allele size. The analysed SSRs and their chromosome arm location were: *Xgwm2* (3AS), *Xgwm11* (1BS), *Xgwm46* (7BS), *Xgwm60* (7AS), *Xgwm88* (6BL), *Xgwm95* (2AS), *Xgwm120* (2BL), *Xgwm136* (1AS), *Xgwm148* (2BS), *Xgwm154* (5AS), *Xgwm155* (3AL), *Xgwm156* (5AL), *Xgwm234* (5BS), *Xgwm251* (4BL), *Xgwm299* (3BL), *Xgwm332* (7AL), *Xgwm389* (3BS), *Xgwm408* (5BL), *Xgwm445* (2AL), *Xgwm513* (4BS), *Xgwm570* (6AL), *Xgwm577* (7BL), *Xgwm601* (4AL). The primer pair for *Xgwm332* amplified alleles from two separate microsatellite loci in all accessions.

Statistical analysis

Similarity and dissimilarity matrices were generated for accessions within duplicates. Three pairwise indexes were calculated: the distance of Rogers (1972) as modified by Wright (1978), $d_{RW,}$ for gliadin and HMWG subunit alleles, Jaccard (1908) similarity index, s_J, for SSRs, and Goldstein and Pollock (1997) distance, d_{GP}, for SSRs. The threshold of genetic difference between accessions in a potential duplicate group was the smallest value observed between any pair of distinct accessions. For heterogeneous accessions at gliadin loci the percent pattern homology (PH) between gliadin banding patterns (Sapirstein and Bushuk 1985) and the Nei (1972) coefficient of genetic identity (I) were computed. Relationships between variables within duplicates were examined by Pearson correlation coefficients. The Goldstein and Pollock distances for SSRs were calculated with the computer package SPAGeDi 1.1 (Hardy and Vekemans, 2002) and the rest of the analyses were performed with the NTSYS-pc software (Rohlf, 1992).

Results

Protein analysis

Gliadin and HMWG subunit alleles of the varieties studied are shown in Table 1. Thirty-nine accessions were characterised by one specific gliadin genotype (gt. 1) and identified as monomorphic. The rest comprised one gliadin more frequent genotype (gt. 1) and one or two less frequent (gt. 2 and gt. 3). All the accessions involved in the same potential duplicate had the same gliadin gt. 1 except for the variety group Mindum and accession 3 of Semental (Table 1). All the gt. 1 identical for gliadins showed no differences in HMWG subunits and total protein, except for Berberisco and Recion duplicates. So, three grains of accession 2 of Berberisco (gt. 1+) differed from gt. 1 in one total protein band. The gt. 1+ of accession 2 of Recion differed from gt. 1 in the slightly lower mobility of the HMWG subunit 8 encoded at *Glu-B1*. Except for Mindum and accession 3 of Semental, d_{RW} between gt. 1 of accessions in the same potential duplicate group was 0. The minimum d_{RW} between known distinct accessions was 0.353. Accessions of Mindum and accession 3 of Semental showed larger values between gt. 1 of their duplicate group ($d_{RW} = 0.707$ for Mindum and 0.866 for Semental). In contrast accession 3 of Semental presented a $d_{RW} = 0$ with gt. 1 of Recio de Baza accessions.

SSRs analysis

Table 2 shows the comparison of gt. 1 for agro-morphological characters and SSRs between accessions

Table 1. Gliadin (*Gli-* alleles), high molecular weight glutenin (HMWG) subunits encoded at *Glu-* loci and total protein of the potential duplicate groups of durum wheat accessions

Variety group	Accession no.	Genotype[a]		*Gli-A1*	*Gli-A3*	*Gli-B5*	*Gli-B1*	*Gli-A2*	*Gli-B2*	*Glu-A1*	*Glu-B1*	Total protein[b]
		No.	Frequency									
Alaga	**1**	**1**	**1.00**	***b***	—	***a***	***new-1***	***k***	***new-1***	***1***	***13+16***	**S**
	2	**1**	**1.00**	***b***	—	***a***	***new-1***	***k***	***new-1***	***1***	***13+16***	**S**
Almendral	**1**	**1**	**1.00**	***b***	***a***	***o***	***b***	***f***	***h***	***N***	***13+16***	**S**
	2	**1**	**0.50**	***b***	***a***	***o***	***b***	***f***	***h***	***N***	***13+16***	**S**
		2	0.30	*b*	*a*	*a*	*new-1*	*f*	*h*	*2**	*13+16*	D
		2+	0.20	*b*	*a*	*a*	*new-1*	*f*	*h*	*N*	*6+8*	D
Andalucía 344	**1**	**1**	**0.70**	***b***	—	***o***	***new-5***	***g***	***h***	***2****	***6+8***	**S**
		2	0.20	*b*	—	*o*	*new-4*	*g*	*h*	*2**	*20x+20y*	D
		3	0.10	*c*	—	*o*	*new-5*	*g*	*h*	*2**	*6+8*	D
Blanquillon de Boñar	**1**	**1**	**1.00**	***b***	—	***a***	***new-1***	***k***	***l***	***2****	*7*	**S**
	2	**1**	**1.00**	***b***	—	***a***	***new-1***	***k***	***l***	***2****	*7*	**S**
Blanco de Corella	**1**	**1**	**1.00**	***b***	—	***a***	***new-1***	***f***	***new-1***	***2****	***6+8***	**S**
	2	**1**	**1.00**	***b***	—	***a***	***new-1***	***f***	***new-1***	***2****	***6+8***	**S**
Berberisco	**1**	**1**	**1.00**	***e***	—	***o***	***c***	***new-1***	***l***	***N***	***20x+20y***	**S**
	2	**1**	**0.30**	***e***	—	***o***	***c***	***new-1***	***l***	***N***	***20x+20y***	**S**
		1+	0.30	*e*	—	*o*	*c*	*new-1*	*l*	*N*	*20x+20y*	D
		2	0.40	*c*	*a*	*o*	*c*	*new-1*	*l*	*N*	*20x+20y*	D
Carita de Ratón	**1**	**1**	**1.00**	***e***	—	***o***	***c***	***new-2***	***t***	***N***	***6+8***	**S**
	2	**1**	**1.00**	***e***	—	***o***	***c***	***new-2***	***t***	***N***	***6+8***	**S**
Fanfarrón	**1**	**1**	**1.00**	***g***	***a***	***o***	***a***	***g***	***new-4***	***N***	***6+8***	**S**
	2	**1**	**1.00**	***g***	***a***	***o***	***a***	***g***	***new-4***	***N***	***6+8***	**S**
Forment	**1**	**1**	**1.00**	***b***	***a***	***a***	***new-2***	***a***	***new-2***	***2****	***6+8***	**S**
	2	**1**	**1.00**	***b***	***a***	***a***	***new-2***	***a***	***new-2***	***2****	***6+8***	**S**
Granja de Badajoz	**1**	**1**	**1.00**	***e***	—	***a***	***c***	***b***	***t***	***N***	***20x+20y***	**S**
	2	**1**	**0.90**	***e***	—	***a***	***c***	***b***	***t***	***N***	***20x+20y***	**S**
		2	0.10	*b*	—	*o*	*b*	*new-3*	*h*	*N*	*6+8*	D
	3	**1**	**0.90**	***e***	—	***a***	***c***	***b***	***t***	***N***	***20x+20y***	**S**
		2	0.10	*c*	—	*a*	*c*	*g*	*h*	*N*	*Het*[b]	D
Hymera	**1**	**1**	**1.00**	***c***	***a***	***o***	***c***	***g***	***t***	***N***	***20x+20y***	**S**
	2	**1**	**1.00**	***c***	***a***	***o***	***c***	***g***	***t***	***N***	***20x+20y***	**S**
Lebrija	**1**	**1**	**0.90**	***b***	***a***	***o***	***b***	***o***	***h***	***N***	***20x+20y***	**S**
		2	0.10	*b*	—	*o*	*new-3*	*o*	*h*	*N*	*6+8*	S
	2	**1**	**0.70**	***b***	***a***	***o***	***b***	***o***	***h***	***N***	***20x+20y***	**S**
		2	0.10	*b*	*a*	*a*	*new-1*	*o*	*h*	*N*	*20x+20y*	D
		2+	0.10	*b*	*a*	*a*	*new-1*	*o*	*h*	*1*	*6+8*	D
		3	0.10	*new-2*	—	*a*	*c*	*o*	*h*	*N*	*20x+20y*	D
Ledesma	**1**	**1**	**1.00**	***c***	—	***o***	***c***	***g***	***h***	***N***	***20x+20y***	**S**
	2	**1**	**1.00**	***c***	—	***o***	***c***	***g***	***h***	***N***	***20x+20y***	**S**
Marqués	**1**	**1**	**1.00**	***c***	—	***o***	***c***	***b***	***new-3***	***N***	***6+8***	**S**
	2	**1**	**0.90**	***c***	—	***o***	***c***	***b***	***new-3***	***N***	***6+8***	**S**
		2	0.10	*c*	*a*	*o*	*c*	*b*	*new-3*	*N*	*6+8*	Het.
Mindum	**1**	**1**	**1.00**	***c***	—	***a***	***a***	***f***	***h***	***N***	***6+8***	**D**
	2	**1**	**1.00**	***a***	—	***a***	***a***	***a***	***a***	***N***	***7+8***	**D**
Raspinegro	**1**	**1**	**1.00**	***c***	***a***	***o***	***c***	***o***	***h***	***N***	***20x+20y***	**S**
	2	**1**	**0.90**	***c***	***a***	***o***	***c***	***o***	***h***	***N***	***20x+20y***	**S**
		2	0.10	*c*	—	*o*	*c*	*o*	*h*	*N*	*20x+20y*	S

Table 1 (cont.). Gliadin (*Gli-* alleles), high molecular weight glutenin (HMWG) subunits encoded at *Glu-* loci and total protein of the potential duplicate groups of durum wheat accessions

Variety group	Accession no.	Genotype[a]		*Gli-A1*	*Gli-A3*	*Gli-B5*	*Gli-B1*	*Gli-A2*	*Gli-B2*	*Glu-A1*	*Glu-B1*	Total protein[b]
		No.	Frequency									
Recio de Baza	**1**	**1**	**1.00**	***b***	—	***o***	***b***	***new-3***	***h***	***N***	***6+8***	**S**
	2	**1**	**1.00**	***b***	—	***o***	***b***	***new-3***	***h***	***N***	***6+8***	**S**
Reción	**1**	**1**	**1.00**	***b***	—	***o***	***c***	***new-1***	***t***	***N***	***6+8***	**S**
	2	**1**	**0.70**	***b***	—	***o***	***c***	***new-1***	***t***	***N***	***6+8***	**S**
	2	1+	0.30	*b*	—	*o*	*c*	*new-1*	*t*	*N*	*6+8'*	S
Rubio de Badajoz	**1**	**1**	**1.00**	***e***	—	***a***	***b***	***a***	***new-5***	***N***	***6+8***	**S**
	3	**1**	**1.00**	***e***	—	***a***	***b***	***a***	***new-5***	***N***	***6+8***	**S**
Rubio de Miajadas	**1**	**1**	**1.00**	***b***	***a***	***a***	***c***	***new-4***	***h***	***N***	***6+8***	**S**
	2	**1**	**1.00**	***b***	***a***	***a***	***c***	***new-4***	***h***	***N***	***6+8***	**S**
Semental	**1**	**1**	**1.00**	***new-1***	—	***o***	***a***	***b***	***a***	***1***	***20x+20y***	**S**
	2	**1**	**0.90**	***new-1***	—	***o***	***a***	***b***	***a***	***1***	***20x+20y***	**S**
		2	0.10	*Het.*	—	*o*	*a*	*b*	*new-6*	*1*	*20x+20y*	D
	3	**1**	**1.00**	***b***	—	***o***	***b***	***new-3***	***h***	***N***	***6+8***	**D**
Senatore Capelli	**1**	**1**	**1.00**	***c***	***a***	***o***	***c***	***o***	***h***	***N***	***20x+20y***	**S**
	2	**1**	**1.00**	***c***	***a***	***o***	***c***	***o***	***h***	***N***	***20x+20y***	**S**
	3	**1**	**1.00**	***c***	***a***	***o***	***c***	***o***	***h***	***N***	***20x+20y***	**S**
Verdial	**1**	**1**	**1.00**	***e***	***a***	***o***	***c***	***new-4***	***l***	***2****	***6+8***	**S**
	2	**1**	**1.00**	***e***	***a***	***o***	***c***	***new-4***	***l***	***2****	***6+8***	**S**

[a] The main genotype 1 of each accession is in bold. [b] S, similar; D, dissimilar, to the gt. 1 of the duplicate. Het: heterozygous.

within potential duplicate groups. Nine out of the 23 potential duplicates showed no agro-morphological differences between duplicate accessions. The rest were distinct in up to four agro-morphological characters (Table 2). Differences in SSRs between gt. 1 of duplicates ranged from 0 to 19 alleles. Most of them differed in two or fewer SSRs. The maximum s_J for SSRs between distinct accessions was 0.777. The two accessions of Alaga, Berberisco, Forment, Mindum and accession 3 of Semental displayed shorter similarity values within duplicates. However, accessions in Alaga, Berberisco, and Forment groups were closer to each other than to other accessions. They differed in 4 SSRs at least, and in 1 or 2 agro-morphological characters. The cut off threshold for d_{GP} was 50.291. The two accessions of Alaga, Hymera, Fanfarron, Mindum, accession 2 of Rubio de Badajoz and accession 3 of Semental showed larger distances within duplicates, being closely related to other accessions. Differences within these duplicates varied from 1 to 19 SSRs and from 0 to 4 agro-morphological characters (Table 2). An inter-group duplication was detected between Raspinegro and 'Senatore Capelli' for gliadins, HMWG subunits, and total protein (Table 1). The duplication was confirmed with SSRs ($s_J \geq 0.777$ and $d_{GP} \leq 4.333$) and with the agro-morphological data.

Intra-accession variability

For heterogeneous accessions several genetic similarity parameters were analysed to compare the intra-accession genotypes, gt. 1+, gt. 2, gt. 2+, and gt. 3, with gt. 1 (Table 3). Differences involved 0 to 5 gliadin loci and 2 to 16 SSRs. Dissimilarities in two or more gliadin loci (putative off-types) were analysed separately from differences in one or no loci (probable biotypes). Protein analysis indicated that 11 genotypes were probable off-types differing in two gliadin loci at least (Table 1). All of them were distinct in HMWG subunits and/or total protein. These genotypes had $I \leq 0.66$ and $PH < 80\%$ with gt. 1, except for gt. 2 of accession 1 of Lebrija $PH = 81.81$ (Table 3). Genetic distances d_{RW} were shorter with other accessions than with accessions included in the same duplicate. The same results were obtained with s_J for SSRs data although gt. 2 of accession 3 of Granja de Badajoz was not analysed. All these possible off-types were confirmed with d_{GP}, except for

Table 2. Comparison of the genotypes 1 for agro-morphological characters and SSRs between accessions of the potential duplicate groups

Variety group	Accession compared	Unshared agro-morphological characters	Unshared SSRs	s_J[a]	d_{GP}[b]
Alaga	1 *vs* 2	Grain color, Days to flower	7	0.548	86.208
Almendral	1 *vs* 2	None	0	1.000	0.000
Blanquillon de Boñar	1 *vs* 2	None	0	1.000	0.000
Blanco de Corella	1 *vs* 2	Flag leaf habit	0	1.000	0.000
Berberisco	1 *vs* 2	Mature spike habit	4	0.714	1.166
Carita de Ratón	1 *vs* 2	Flag leaf habit	1	0.920	0.166
Fanfarrón	1 *vs* 2	None	2	0.846	5,704.335
Forment	1 *vs* 2	Glume-shoulder shape, Glume-shoulder length	4	0.714	12.083
Granja de Badajoz	1 *vs* 2	Flag leaf habit	0	1.000	0.000
	3 *vs* 1, 2	None	1	0.920	0.166
Hymera	1 *vs* 2	Anthers pigment	1	0.920	3,750.000
Lebrija	1 *vs* 2	Flag leaf habit, Anthers pigment	0	1.000	0.000
Ledesma	1 *vs* 2	None	0	1.000	0.000
Marqués	1 *vs* 2	Days to flower	0	1.000	0.000
Mindum	1 *vs* 2	Awn color, Plant height	19	0.116	951.416
Raspinegro	1 *vs* 2	Flag leaf habit	0	1.000	0.000
Recio de Baza	1 *vs* 2	None	0	1.000	0.000
Reción	1 *vs* 2	None	0	1.000	0.000
Rubio de Badajoz	1 *vs* 3	None	0	1.000	0.000
	2 *vs* 1, 3	Mature spike habit	1	0.920	5,766.000
Rubio de Miajadas	1 *vs* 2	Flag leaf habit	1	0.920	0.166
Semental	1 *vs* 2	Flag leaf habit, Mature spike habit	0	1.000	0.000
	3 *vs* 1, 2	Glume hairiness, Glume colour, Glume-beak curvature, Glume-beak length	16	0.200	1,820.958
Senatore Capelli	3 *vs* 1, 2	Mature spike habit, Days to maturity	1	0.920	1.041
	1 *vs* 2, 3	Glume length, Plant height	2	0.846	6.208
Torcal	1 *vs* 2	None	3	0.777	0.500
Verdial	1 *vs* 2	Awn colour, Glume internal hairs, Glume-beak length	3	0.777	16.708

[a] Jaccard similarity index. [b] Goldstein and Pollock distance.

gt. 2 of accession 1, and gt. 2 and gt. 3 of accession 2 of Lebrija group.

Analysis of potential biotypes indicated that four genotypes (in Andalucia 344, Marques and Raspinegro accessions) differed from gt. 1 in one gliadin locus (Table 1). These genotypes possessed I > 83% and PH > 80% with gt. 1. Except for Raspinegro, all of them were distinct from gt.1 in total protein (Tables 1 and 3). A more precise analysis of the Marqués accession indicated that gt. 2 was a heterozygote since total protein gt. 1 was included in gt. 2. Genotype 2 of Andalucia 344 also diverged in HMWG subunits from gt. 1. The values of d_{RW} with gt. 1 were shorter than that found between different accessions, except for gt. 2 of Andalucia 344. SSR data indicated that these potential biotypes differed from gt. 1 in 9 to 12 alleles and overtook the threshold established for s_J and d_{GP}. However, they were more related to gt. 1 than to other accessions based on s_J values. In two additional cases intra-accession genotypes were identical in gliadins but showed slight differences in HMWG subunits or total protein (gt. 1+ of accession 2 of Berberisco and gt. 1+ of accession 2 of Recion, Table 1). Both were closely related to gt. 1 with d_{RW} (Table 3). For SSRs, they were also grouped with s_J but not with d_{GP}. In only one case an intra-accession genotype possessed the same protein composition as the gt.1 of a distinct duplicate group: gt. 2 of accession 2 of Granja de Badajoz with gt. 1 of Recio de Baza (Table 1). They differed in 3 SSRs and were grouped with both s_J and d_{GP} (Table 3).

Pearson correlation coefficients (data not shown) within duplicates indicated that the number of unshared SSRs, and s_J, were significantly correlated with the number of unshared agro-morphological traits ($p < 0.01$), PH ($p < 0.05$), and I ($p < 0.01$), while d_{GP} was not significantly correlated with the three later

Table 3. Comparison of genotypes (gt.) of heterogeneous accessions with the main genotype (gt. 1) for gliadin, high molecular weight glutenin (HMWG) subunits and total protein.

Variety	Accession number	gt. compared with gt. 1	Unshared gliadin loci	%PH[a]	I[b]	HMWG/ total protein[c]	d_{RW}[d]	Unshared SSRs	s_J[e]	d_{GP}[f]
Almendral	2	2	2	73.91	0.66	D/D	0.612	14	0.263	14,852.875
	2	2+	2	73.91	0.66	D/D	0.612	12	0.333	4,371.833
Andalucía 344	1	2	1	82.35	0.83	D/D	0.500	12	0.333	4,656.583
	1	3	1	88.23	0.83	S/D	0.353	9	0.454	11,046.667
Berberiso	2	1+	0	100.00	100.00	S/D	0.000	2	0.846	1,697.708
	2	2	2	69.23	0.66	S/D	0.500	6	0.600	117.208
Granja de Badajoz	2	2	5	18.52	0.16	D/D	0.866	16	0.200	1,899.000
	3	2	3	28.12	0.50	D/D	0.661	—	—	—
G. de Badajoz[g]	2	2	0	100.00	100.00	S/S	0.000	3	0.777	40.375
Lebrija	1	2	2	81.81	0.66	D/S	0.612	9	0.454	40.958
	2	2	2	69.56	0.66	S/D	0.500	6	0.600	40.166
	2	2+	2	69.56	0.66	D/D	0.707	8	0.500	331.791
	2	3	4	57.69	0.33	S/D	0.707	5	0.655	18.500
Marqués	2	2	1	83.33	0.83	S/D	0.353	10	0.411	7,750.666
Raspinegro	2	2	1	91.67	0.83	S/S	0.353	—	—	—
Reción	2	1+	0	100.00	100.00	D/S	0.353	2	0.846	60.333
Semental	2	2	2	65.22	0.66	S/D	0.500	8	0.500	415.041

[a] Percent gliadin pattern homology. [b] Nei's coefficient of genetic identity based on gliadin alleles. [c] S,D are similar and dissimilar, respectively, to the gt. 1 of the duplicate. [d] Distance of Rogers as modified by Wright based on gliadin and HMWG subunits alleles. [e] Jaccard similarity index based on SSRs alleles. [f] Goldstein and Pollock distance based on SSRs alleles. [g] The compared gt. 1 was from Recio de Baza.

variables. Correlations were significant ($p < 0.01$) between PH and I, and between s_J and d_{GP}.

Discussion

Most of the potential duplicate groups analysed in this study refer to accessions derived from the same original population collected at the beginning of the 20th century. These accessions were maintained in different breeder collections, or came from the USDA-ARS germplasm collection. To characterise probable duplication, eight protein loci (coding for gliadins and HMWG subunits) and 24 SSRs were used. This number is comparable with that reported by Virk *et al.* (1995) who demonstrated that 26 polymorphic markers were enough to detect at least one difference between suspected pairs of rice duplicates at 99% probability. The 24 SSRs were selected based on their polymorphism in wheat and their position on the genetic map (Röder *et al.,* 1998; Maccaferri *et al.,* 2003). Except for 1AL, 1BL, 4AS, 6AS, and 6BS, all the chromosome arms were covered. Nevertheless, 6AS and 6BS were analysed with the gliadin loci *Gli-2*, and 1AL and 1BL with HMWG subunit loci *Glu-1*. Moreover, the information from total protein analyses allowed coverage of a wider genome area.

Twenty-one out of the 23 potential duplicates analysed possessed the same gt. 1 for gliadins, HMWG subunits and total protein (Table 1). They differed in three agro-morphological characters at most. In general, these characters had low discriminating power (flag leaf habit, mature spike habit, or glume length) or were affected by environmental conditions, such as anther pigment (Ruiz and Aguiriano, 2004). The two cases of dissimilarity, Mindum and accession 3 of Semental, were also confirmed with SSRs (Table 2). Semental presented differences in four agro-morphological characters, but Mindum differed only in two traits. So, the later duplicate group was difficult to solve based only on agro-morphological data.

Eighteen of the 21 possible duplicates with identical gt. 1 for proteins showed none or little differences in SSR alleles (from 1 to 3). In agreement with passport

data information, all of them presented shorter distances with accessions of the same potential duplicate than those obtained for distinct accessions. Six of them were identical duplicates with no differences between duplicate accessions for any of the genetic markers analysed (agro-morphological, proteins, and SSRs). Identical duplicates in *ex-situ* collections are not frequent, even in an autogamous species like wheat.

In three cases, discrepancies in variation detected by proteins and SSRs, and also between s_J and d_{GP} indexes, were found. Three duplicates were verified with s_J (Hymera, Fanfarron, and accession 2 of Rubio de Badajoz) but not with d_{GP}. The discrepancies between both indexes were caused by the presence of null alleles, which increased d_{GP} values. On the contrary, two potential duplicates (Berberisco and Forment) with four unshared SSRs were only verified with d_{GP} because the differences between allele sizes were short. A similar result was obtained with accession 3 of Semental that possessed the same protein composition (Table 1) and morphotype as the two accessions of Recio de Baza. They were separated with s_J (6 unshared SSRs) but not with d_{GP} ($d_{GP} = 0.833$).

The two duplicate accessions of Alaga, with seven unshared SSRs, exceeded the threshold established for s_J and d_{GP}. However, the two accessions were closer related to each other than to other accessions with s_J. Both accessions seem to be agro-types with contrasting seed colour and days to flower (Table 2). In agreement with Aguiriano *et al*. (2006), in the past, it could be intentional splitting of the original sample into morphologically distinct parts. Tranquilli *et al*. (2000) also found an Argentinean wheat landrace with considerable variation in agro-morphological characters but no variation in HMWG subunits and isozymes. The diversity observed was explained by artificial selection carried out in the local area.

The significant correlations detected between the number of unshared agro-morphological traits and SSRs indicated a good concordance between molecular marker diversity and morphotype when a small number of genotypes are screened (Crouch *et al.,* 2000). However, differences in agro-morphological traits did not imply differences in molecular data and viceversa.

No obvious intra-accession differences for agro-morphological descriptors were detected in heterogeneous accessions at gliadin loci. A gliadin genotype was classified as a biotype if its frequency was larger than 5% and it was different from gt. 1 in not more than one gliadin-coding locus (Metakovsky, 1991). Otherwise, a genotype was classified as an off-type. All the potential off-types for gliadins were confirmed with s_J. They differed in 5-16 SSRs from gt. 1 (Table 3) in agreement with the upper limit of 3 unshared SSRs obtained between gt. 1 of accessions of the same duplicate (Table 2). The potential gliadin off-types were also confirmed with d_{GP} except for Lebrija duplicate. In this case, the genotypes have probably varied for a long time but maintaining a genetic relation with gt. 1. The heterozygosity at two SSR loci in gt. 2 of accession 1 (Table 3) also suggests the possible occurrence of outbreeding during multiplication. Accordingly, safety measures such as species alternation are recommended during regeneration since other cases of heterozygosity were detected (Table 1).

Most of the admixtures observed were heterogeneous at *Gli-B1* (Table 1). It is known that gliadin alleles encoded at this locus have contrasting relationships with quality. Therefore, quality of a sample of a heterogeneous variety may depend on the presence and frequency of diverse gliadin biotypes. In this study, gt. 2 and gt. 2+ of accession 2 of Almendral possessed allele *new-1*, which conferred poorer quality than *b* present in gt. 1 (Aguiriano *et al.,* 2009). This valuable information should be considered in evaluation and utilization of the sample.

The potential gliadin biotypes dissimilar in one locus were not confirmed as biotypes with HMWG subunits or total protein, and SSRs (Table 3). Probably, these closely related genotypes have diversified during maintenance in the Bank. Only those genotypes identical in gliadin composition and showing few differences in HMWG subunits or total protein were verified as biotypes with SSRs: gt. 1+ of accession 2 of Berberisco, gt. 1+ of accession 2 of Reción, and gt. 2 of accession 2 of Granja de Badajoz in comparison with Recio de Baza. The analysis indicated that the biotypes verified with s_J differed in no more than one protein-coding locus whether it coded for gliadins, HMWG subunits, or one band of total protein.

Comparison analyses of the two similarity indexes used to examine gliadin genotypes (PH and I) with SSRs showed that, in general, PH and I with values lower than 80 identified clearly off-types with more than 3 SSRs unshared. Sapirstein and Bushuk (1985) also considered a cutoff threshold of 80 for PH to differentiate varieties. Although band treatment of the information (PH) appears less discriminating, because of redundancy, than allelic treatment (I) both were significantly correlated. However, correlations with the number

of unshared SSRs were higher with I than with PH. One disadvantage of the analysis of PH over allelic composition is the difficulty of comparing protein patterns of different varieties evaluated in different electrophoretic gels. The mistakes in accession 3 of Semental and gt. 2 of accession 2 of Granja de Badajoz (both similar to Recio de Baza), and the duplication between Raspinegro and Senatore Capelli, were detected based on protein alleles and confirmed with SSRs. In contrast, identification of alleles is sometimes complicated and labour-consuming.

In the present research s_J and d_{GP} were significantly correlated. Jaccard's index is based on allele frequencies shared and an infinite allele mutation model. Goldstein and Pollock distance, was developed specifically for microsatellite applications and assumes a single-step mutation model. With Jaccard's no homoplasy exists and it can overestimate similarities among individuals. This overestimation was not observed in this study. Possible bias to homoplasy could be minimized by the use of dinucleotide loci and compound repeats as suggested Maccaferri *et al.* (2003). The discrepancies between both genetic indexes were mainly due to the incidence of null alleles. Moreover, similarities between and within accessions not found with s_J were detected with d_{GP}. In these cases, allele differences were high in number but not in size, indicating that there was a close relation between the genotypes compared. From the standpoint of managing large *ex-situ* germplasm collections, the fate of not identical but closely related accessions and genotypes are complex decisions.

In general, a good concordance among proteins, morphological traits and SSR were detected, mainly when the differences were high. However, the discrepancy in similarity detected in some cases between the different data sets showed that it is advisable to evaluate germplasm through distinct approaches. The three types of proteins used were an excellent tool to detect duplicates and discriminate among admixtures in the accessions. In addition, protein analysis has a lower cost (economic and personnel) than agro-morphological and SSRs data. There was a good concordance between the two SSRs indexes used. In the cases of discrepancy both reported complementary information if null alleles are avoided. The analysis together of proteins and SSRs data are very useful to identify potential duplicates, mistakes not found with passport information and to distinguish authentic biotypes, closely related genotypes, and contaminations.

Acknowledgments

This work was supported by grant RFP2008-00006-00-00 from *Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria* (INIA) of Spain and AGL 2006-8025 from *Comisión Interministerial de Ciencia y Tecnología* (CICYT) of Spain. E. Aguiriano has received a pre-doctoral scholarship from the INIA.

References

AGUIRIANO E., RUIZ M., FITÉ R., CARRILLO J.M., 2006. Analysis of genetic variability in a sample of the durum wheat Spanish collection based on gliadin markers. Genet Resour Crop Ev 53, 1543-1552.

AGUIRIANO E., RUIZ M., FITÉ R., CARRILLO J.M., 2009. Effects of N fertilisation, year and prolamin alleles on gluten quality in durum wheat (*Triticum turgidum* L. ssp. *turgidum*) landraces from Spain. Span J Agric Res 7, 342-348.

CROUCH K., CROUCH J.H., MADSEN S., VUYLSTEKE D.R., ORTIZ R., 2000. Comparative analysis of phenotypic and genotypic diversity among plantain landraces (*Musa* spp., AAB group).Theor Appl Genet 101, 1056-1065.

DOYLE J.J., DOYLE J.L., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13-15.

FRA-MON P., SALCEDO G., ARAGONCILLO C., GARCÍA-OLMEDO F., 1984. Chromosomal assignment of genes controlling salt-soluble proteins (albumins and globulins) in wheat and related species. Theor Appl Genet 69, 167-172.

GHOSH S., KARANJAWALA Z.E., HAUSER E.R., ALLY D., KNAPP J.I., RAYMAN J.B., MUSICK A., TANNENBAUM J., TE C., SHAPIRO S., 1997. Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. Genome Res 7, 165-178.

GOLDSTEIN D.B., POLLOCK D.D., 1997. Launching microsatellites: a review of mutation processes and method for phylogenetic inference. Genetics 139, 463-471.

HARDY O.J., VEKEMANS X., 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2, 618-620.

HINTUM TH.J.L. VAN, KNUPFFER H., 1995. Duplication within and between germplasm collections. I. Identifying duplication on the basis of passport data. Genet Resour Crop Evol 42, 127-133.

JACCARD P., 1908. Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44, 223-270. [In French].

KUDRYAVTSEV A.M., BOGGINI G., BENEDETTELLI S., ILLICHEVSKII N.N., 1996. Gliadin polymorphism and genetic diversity of modern Italian durum wheat. J Genet Breed 50, 239-248.

LAFIANDRA D., KASARDA D.D., 1985. One- and two-dimensional (two pH) polyacrylamide gel electrophoresis in a single gel: separation of wheat proteins. Cereal Chem 62, 314-319.

LUND B., ORTIZ R., SKOVGAARD I.M., WAUGH R., ANDERSEN S.B., 2003. Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data. Theor Appl Genet 106, 1129-1138.

MACCAFERRI M., SANGUINETI M.C., DONINI P., TUBEROSA R., 2003. Microsatellite analysis reveals a progresive widening of the genetic basis in the elite durum wheat germplasm. Theor Appl Genet 107, 783-797.

METAKOVSKY E.V., 1991. The value of gliadin biotypes in commercial cultivars of wheat. In: Gluten proteins 1990 (Bushuk W., Tkachuk R., eds). AACC, St Paul, Mi, USA. pp. 569-580.

NEI M., 1972. Genetic distance between populations. Am Nat 106, 283-292.

PAYNE P.I., HOLT L.M., LAWRENCE C.J., LAW C.N., 1982. The genetics of gliadin and glutenin the major storage proteins of the wheat endosperm. Qualtitas Plantarium-Plant Foods for Human Nutrition 31, 229-241.

PAYNE P.I., LAW C.N., MUDD E.E., 1980. Control of homoeologous group 1 chromosomes of the high molecular-weight subunit, a major protein of wheat endosperm. Theor Appl Genet 58, 113-120.

PAYNE P.I., LAWRENCE C.J., 1983. Catalogue of alleles for the complex gene loci, *Glu-A1*, *Glu-B1*, *Glu-D1* which code for high-molecular-weight subunits of glutenin in hexaploid wheat. Cereal Res Commun 11, 29-35.

RÖDER M.S., KORZUN V., WENDEHAKE K., PLASCHKE J., TIXIER M.H., LEROY P., GANAL M.W., 1998. A microsatellite map of wheat. Genetics 149, 2007-2023.

ROGERS J.S., 1972. Measures of genetic similarity and genetic distance. In: Studies in genetics VII. University of Texas, Austin, Tx, USA. pp.145-153.

ROHLF F.J., 1992. NTSYS-pc Numerical taxonomy and multivariate analysis system. Exeter software, New York.

RUIZ M., AGUIRIANO E., 2004. Analysis of duplication in the Spanish durum wheat collection maintained in the CRF-INIA on the basis of agro-morphological traits and gliadin proteins. Genet Resour Crop Evol 51, 231-235.

SAPIRSTEIN H.D., BUSHUK W., 1985. Computer-aided analysis of gliadin electrophoregrams. II. Wheat cultivar identification and class comparisons. Cereal Chem 62, 377-392.

SINGH J., SKERRITT J.H., 2001. Chromosomal control of albumins and globulins in wheat grain assessed using different fractionation. J Cereal Sci 33, 163-181.

TRANQUILLI G., APPENDINO M.L., PFLÜGER L.A., BULLRICH L., LEWIS S., SUÁREZ E.Y., 2000. Morphoagronomic and biochemical variation in an Argentian landrace of wheat. Genet Resour Crop Evol 47, 281-284.

VIRK P.S., NEWBURY H.J., JACKSON M.T., FORD-LLOYD B.V., 1995. The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theor Appl Genet 90, 1049-1055.

WRIGHT S., 1978. Evolution and the genetics of populations, vol 4. Variability within and among natural populations. University of Chicago Press, Chicago.